

SPECIAL ISSUE PAPER

Classification of human motion based on affective state descriptors[†]

Gokcen Cimen^{1*}, Hacer Ilhan², Tolga Capin¹ and Hasmet Gurcay²¹ Bilkent University, Computer Engineering Department, Ankara, Turkey² Hacettepe University, Department of Mathematics, Ankara, Turkey

ABSTRACT

Human body movements and postures carry emotion-specific information. On the basis of this motivation, the objective of this study is to analyze this information in the spatial and temporal structure of the motion capture data and extract features that are indicative of certain emotions in terms of affective state descriptors. Our contribution comprises identifying the directly or indirectly related descriptors to emotion classification in human motion and conducting a comprehensive analysis of these descriptors (features) that fall into three different categories: posture descriptors, dynamic descriptors, and frequency-based descriptors in order to measure their performance with respect to predicting the affective state of an input motion. The classification results demonstrate that no single category is sufficient by itself; the best prediction performance is achieved when all categories are combined. Copyright © 2013 John Wiley & Sons, Ltd.

KEYWORDS

motion capture; human motion classification; affective state; emotion

*Correspondence

Gokcen Cimen, Bilkent University, Ankara, Turkey.

E-mail: gokcen.cimen@cs.bilkent.edu.tr

1. INTRODUCTION

Much of the difficulty in creating character animation is to give the character representation of expression or personality. Even though, in traditional animation, animators can achieve this challenge in a time-consuming key-framing processes, the spreading availability of motion capture technology and large amount of human motion databases have increased the importance of reusability of motion capture data in human animation based on the analysis and understanding of human motion. It is well known that the personality of a character is conveyed through emotions, and human motion carries important visual cues associated with emotions. On the basis of this motivation, the objective of this study is to analyze these cues from spatial and temporal structure of the motion capture data and classify human motions according to one of their stylistic variations—emotion or the energy level of the actor (low energy interpreted as sad, relaxed; high energy as happy, angry).

The term *emotion* is sometimes used in reference to the emotional state and is also called the *affective state*, which

has been studied in detail in emotion research and affective science. Affective science is mainly concerned with the questions of which features of human motion are important for emotional communication and how human beings use them in order to distinguish one emotion from another. Dimensional models of emotion or affective state can lie in two or three dimensions. Several two-dimensional models of emotion have been developed. In our study, among these models, we apply the circumplex model of emotion that was first developed by James Russell [1]. Figure 1 shows the circumplex model of affective states. In this context, we will use the terms emotion and affective state interchangeably.

Although it has been debated that the specific body movements may be only indicative of the intensity of the emotion, our study presents an affective state analysis and classification of human motion capture data, which attempts to demonstrate that emotion-specific cues in body movements and postures are indicative of certain emotions. Our main contributions comprise defining and categorizing the affective state descriptors that are mostly adapted from the past and current approaches in affective science and attempting to identify the directly or indirectly related descriptors to emotion classification in human motion. We have investigated three categories of descriptors—posture-based, dynamic, and frequency-based—and measured their

[†]Supporting information may be found in the online version of this article.

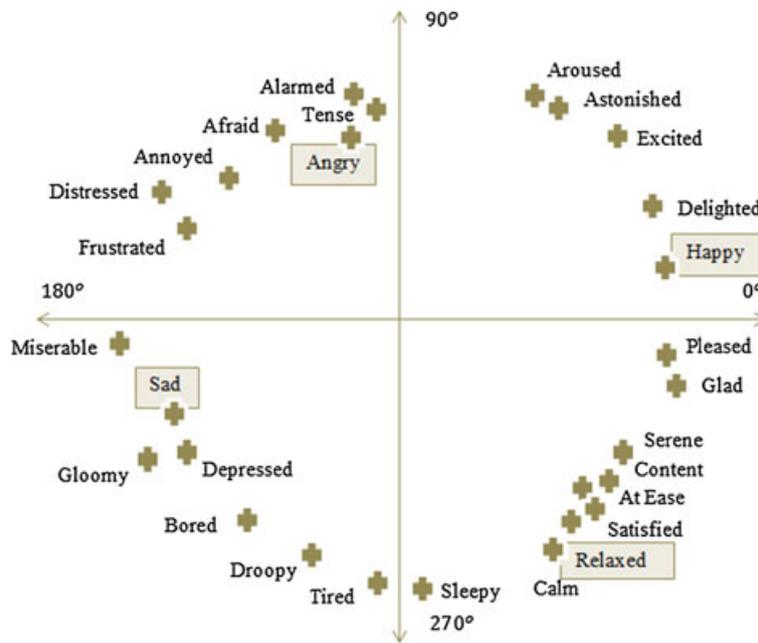


Figure 1. The circumplex model (Russell 1980).

performance with respect to predicting the affective state of the input motion. The results indicate that no single category is sufficient to predict the emotional state by itself, and the best prediction performance is achieved when all the categories are combined.

2. RELATED WORK

The motion recognition and classification problem draws interest in a variety of major disciplines including robotics, computer animation, and psychology. It comprises identifying two different types of human actions—primary and secondary themes [2]. Whereas primary themes represent basic human actions such as walking, running, or jumping, secondary themes consider the stylistic variations in the motion caused by the actor's emotion, gender, or age. Our approach falls into the second type of classification with respect to variations caused by different actors.

A large body of work exists for the recognition and classification of human action based on the primary theme. These approaches aim the recognition of primary human actions such as walking, running, kicking, and punching and classifying them according to their spatial and temporal structure. Among them, a considerable amount of research focus on the recognition of human actions from video images [3,4]. An alternative to detecting actions from video sequences is a data-driven approach, using available 3D motion capture technology [5,6].

In these works, the motion recognition and classification methods have been based on primary human actions regardless of the actor's style. Our goal is to analyze and classify human motions associated with specific

emotions, which falls into the secondary theme. Related to our work, a number of studies target the classification of the emotional state in human motion. Troje has applied [7] Principal Component Analysis (PCA) in order to learn lower-dimensional representations of human walking in different emotional states. There have also been studies detecting emotion from static body posture descriptors. Coulson has conducted experiments [8] where observers classified the emotions from static images. Another important descriptor for the recognition of emotion in motion is kinematical information, such as velocity, acceleration, or jerk. Wallbot [9] found that spatio-temporal features of body movements, rather than static body postures are more indicative of differences between emotions and changes in energy states. Atkinson *et al.* [10] investigated the exaggeration of whole body movements and found that exaggeration of body movements would increase the emotional intensity and enhance emotion recognition. In another work, Kapur *et al.* [11] used dynamics of motion features and analyzed the performance of different classifiers including neural networks and support vector machines (SVMs). Although classifiers achieved a correct recognition rate of 84%, in the user study, humans were able to classify the motions with 93% of accuracy.

Although the main aim of this study is the classification of human action and variations, there are also publications that study how to synthesize human motion by understanding and computationally modeling them. Rose *et al.* present the Verbs-and-Adverbs approach for interpolating styles for motion synthesis [12]. Grochow *et al.* use the Scaled Gaussian Process Latent Variable Model [13], which make use of a training set to capture the style of the actor for motion synthesis and inverse kinematics.

3. APPROACH

This paper proposes a style-based classification solution for recognizing and classifying secondary themes in a given motion sequence, on the basis of affective state descriptors. Our supervised motion classification system consists of two phases, as illustrated in Figure 2. The first one is the training phase, where we learn affective state descriptors from a human motion training set. The training set is recorded with multiple actors performing the same walking action in a variety of affective states (sad, happy, angry, and relaxed). Consecutively, we perform space warping and time warping of the input motions to the captured motion sequences in order to bring the input motion data in a form that is invariant with respect to space and time. Then, we learn affective state descriptors from the processed motion data set. For this purpose, we investigate and compare several alternatives for mapping the motion attributes to affective descriptor features.

Then, in the second—prediction—phase, given a new input motion sequence, we classify it with respect to the learned affective states, by utilizing the SVM learning algorithm for classifying the features.

3.1. Data Acquisition

We recorded the training set using a Vicon motion capture system with eight cameras, with capture rate at 120 Hz. Overall, 42 markers were positioned on the suit of the actors following the Vicon motion capture guidelines. The absolute position of the root node and joint orientations in Euler angles were used for representing each pose. The

joints used during the analysis were head, neck, lower back, thorax, right and left collar, humerus, wrist elbow, femur, tibia, and foot.

We recorded the training set using 12 actors, each performing the same action in a variety of emotions (sad, happy, angry, and relaxed). Overall, 96 motion samples were captured. During the motion capture process, all actors were recorded while walking in an approximately straight line with the same number of steps. Each actor performed walking in four different emotional states, each of which was repeated several times. The actors were instructed to avoid the use of secondary gestures that would interrupt their rhythmic walking pattern. For example motion clips used as the training data, please see Supporting information.

3.2. Data Preprocessing

One of the challenges that emerge while performing statistical analysis on time series data such as motion capture sequences is the spatial and temporal misalignments. Therefore, all motion data should be unified in both time and space dimension in order to cancel the variations in spatio-temporal scale of motion examples and to establish the correspondence between different motion sequences. For this purpose, we apply two preprocessing steps: space warping for making the captured motions translation and rotation invariant and time warping for warping the motion samples to the same duration.

Space Warping. The major challenge while analyzing motion data is the variances with respect to the global root

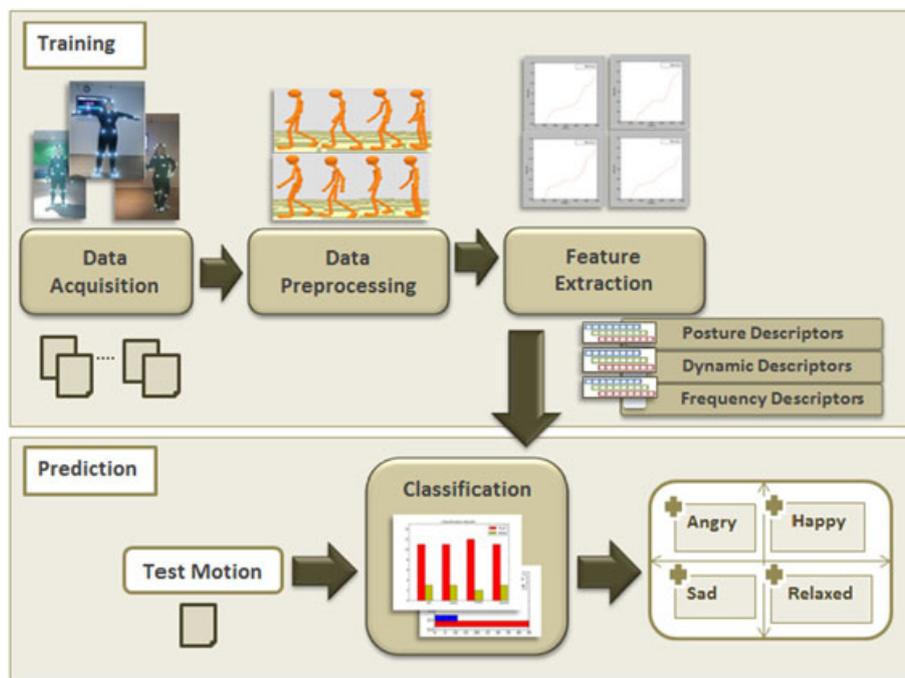


Figure 2. Overview of the system.

position and orientation in the absolute world coordinate system. Therefore, before applying any similarity analysis over the motion data using global positions of the joints, the motion examples should be translation and rotation invariant. We unify the movement direction of each motion sequence, by first translating the motion samples to the origin. Then, to unify their directions, we rotate each motion sequence around the vertical axis so that the direction of each motion is aligned to the same axis (x -axis, in our system). For finding the angle θ for each motion that aligns it to the axis, we use a simple quadratic optimization, similar to the approach of Ma *et al.* [14]:

$$\theta^* = \arg \min_{\theta} \|R(\theta)\text{dir}(M) - x\|^2 \quad (1)$$

Time Warping. In the next step, we apply the dynamic time warping algorithm (DTW) [15] to remove the variations in the dimensions of motion sequences by comparing two sequences of different length for similarity and nonlinearly shifting in time axis. Given two time series $A = (a_1, a_2, \dots, a_n)$ and $B = (b_1, b_2, \dots, b_m)$ in different number of frames, we select the longest sequence as the reference sequence, and the other one is time-warped over the same dimension of the reference sequence. The algorithm firstly constructs the cost matrix by calculating pairwise distance between two sequence, $C \in \mathbb{R}^{n \times m}$. Then, an optimal path is found by walking through the low cost areas in the cost matrix that gives the correspondences between the two sequences. Figure 3 shows the results of two different short motion sequences after space and time warping.

It is well understood that the spatio-temporal characteristics of human motion are good indicators of style. Nonlinear time alignment methods, such as DTW, may change the characterization of temporal stylistic features of the motion [16]. In order not to lose temporal information extracted from dynamic descriptors, we apply DTW to dynamic feature set after the features are extracted.

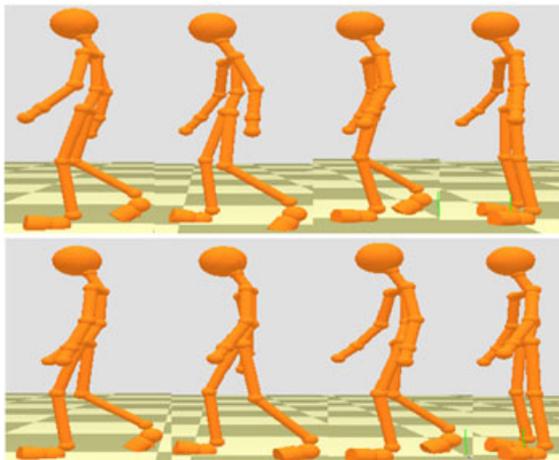


Figure 3. Two different short motion sequences after space and time warping.

3.3. Style-based Feature Extraction

Although the human body can perform complex actions, it is well understood that human motion can also carry the indicative personal information such as the inner intention of the movement or emotional state of the performer. In most situations, these emotion-specific subtle changes in the qualities of movement can be identified and used as the emotional state features.

Our work is mainly concerned with the question of which features of human motion can be used for classification. Therefore, we consider and compare several alternatives of features extracted from motion sequences and represent them as affective state descriptors. It is possible to divide these emotional state descriptors or features into different categories: (i) posture descriptors, (ii) dynamic descriptors, and (iii) frequency-based descriptors. Posture descriptors refer to the features specified in terms of body posture, whereas dynamic descriptors describes the features influenced by movement. Frequency-based descriptors are based on frequency analysis of the motion and can give specific patterns related to emotional-specific cues in the motion. In the next subsections, we first review these descriptors and discuss several alternative forms of setting up feature vectors for classification.

3.3.1. Posture Descriptors.

Posture descriptors are used in emotion recognition and can be extracted from body postures.

End Effector Positions. Although the position of each joint can be selected for the feature vector at first sight, this is not an effective solution. Human body motion is known to have significant redundancies. Furthermore, the importance of each joint may differ from one motion to another. For example, the position of the femur joint gives more information compared with the position of a toe joint in a walking motion.

To overcome this problem, while also avoiding the manual adjustment of joint importance values by the animator, we use the position of the end effectors as positional features in this work. Therefore, we construct a feature vector that consists of four end effectors for the wrist and the ankle joints and an additional end effector position for the head.

The idea of considering the positions of four end effectors and head as a feature vector has already been proposed by Kruger *et al.* [17]. As it is also observed in this work, the positions and the orientations of the end effectors are suitable for describing a human pose because they entirely specify the geometry of the arms and legs.

The end effector position feature vector E is constructed as follows:

$$E = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \quad (2)$$

where n is the number of frames, and vector $\mathbf{e}_i = [p_{x1}, p_{x2}, \dots, p_{x5}]^T$ holds the x -axis component of the five end effector positions. The main reason for excluding the y -axis and z -axis components is that x -axis contains more variance and thus is more suitable for observing the cyclic swinging of the limbs during the motion since the motion directions are aligned to the x -axis in the preprocessing step.

End Effector Orientations. The orientation of the end effectors is also an indicator of postural variations in a motion, and they are statistically dependent on the end effector positions. End effector angles are calculated as the angle θ between successive two segments connected to the end effectors, such as the tibia and toe corresponds the angle of ankle. The output feature vector is thus constructed as follows:

$$T = \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \\ \vdots \\ \mathbf{t}_n \end{bmatrix} \quad (3)$$

where the vector $\mathbf{t}_i = [\theta_1, \theta_2, \dots, \theta_5]^T$ holds the x -axis components of the end effector orientations.

Bounding Box. Another feature that we consider is the bounding box of the whole body pose and its change in frames in the motion. The main reason to select this feature for emotional descriptor is that it is an indicator of how the actor uses the space during the motion, such that energetic motions use more space by spreading their limbs more than the ones with lower energy. The feature vector B that holds the volume of the bounding box for one motion sample with n number of frames is as follows:

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (4)$$

where b_i is a scalar value and gives the volume of the bounding box for the i th frame of the motion.

Weight Shift (Center of Mass). Weight shifting describes how the body's center of mass moves through space. The sense of weight is an essential component of believable characters. How an actor's weight shifts through space during the action can give information about the style of the actor. Given the mass vector of the joint segments m_j , the center of mass for a posture in the motion segment at time t are given by

$$\mathbf{c}_i = \frac{\sum_{j \in N} m_j p_{ij}}{\sum_{j \in N} m_j} \quad (5)$$

where N is the number of segments in the body. The center of mass feature vector C is constructed as follows:

$$C = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \end{bmatrix} \quad (6)$$

where the vector \mathbf{c}_i holds the coordinates of the center of mass for frame i .

3.3.2. Dynamic Descriptors.

Even though the time series data of poses, as described previously, is an important feature set for emotion recognition in motion, dynamic cues such that velocity, acceleration, and jerk of the joints are also needed in order to better recognize certain emotions. Several related studies [8,18] suggest that significant information was lost in expressed emotions by reducing dynamic aspects of body motion to static postures. Dynamic descriptors express perception of emotions influenced by movement kinematics such as speed, acceleration, and jerk derived from first, second, and third derivatives of the joint positions.

Velocity. When a character is in excited or annoyed state (the first and second quadrants in Russel's circumplex model in Figure 1), the timing of his or her movements will be faster. Conversely, when the character is in the third and fourth quadrants, the movements will be slower. Therefore, the linear velocity is first used for constructing the dynamic features.

The velocity feature vector V is constructed as follows:

$$V = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{bmatrix} \quad (7)$$

where \mathbf{v}_i is a $3N$ dimensional vector that holds the linear velocities of each segment at frame i and is calculated as $\mathbf{v}_i = p_i - p_{i-1}$.

Acceleration. The change in the speed gives more information about the mood of the character. For example, the characters with energy would do sharper movements with changing speed. In addition, the acceleration can give information about weight of the character. For example, the more mass the character has, the longer duration the acceleration is required. Given the velocity vector for a posture \mathbf{v}_i at time t , the acceleration vector for a motion sequences at time t is given by

$$\mathbf{a}_i = \mathbf{v}_i - \mathbf{v}_{i-1} \quad (8)$$

and the acceleration feature vector A is constructed similarly.

Jerk. Jerk is the rate of changes of acceleration or force, and it is captured by taking the derivative of acceleration with respect to time. Jerk is a measure for identifying

changes in acceleration and therefore changes in applied torques to the joints. Given the acceleration vector for a posture \mathbf{a}_i at time t , the jerk vector for a motion sequences at time t are given by

$$\mathbf{j}_i = \mathbf{a}_i - \mathbf{a}_{i-1} \tag{9}$$

and the jerk feature vector J is constructed similarly.

3.3.3. Frequency-based Descriptors.

In addition to static and dynamic descriptors listed previously, our third set of descriptors represents the motion data in frequency domain. This provides the ability to determine small signals buried within complex time domain motion signals that can carry information about some characteristic patterns related to emotional cues. When converting signals from time domain to frequency domain, we use the standard fast Fourier transformation (FFT). FFT enable us to decompose signal into components of different frequencies. The input of the FFT is the end effector position vector described in Section 3.3.1, and the output vector contains the sum of the positive and negative-frequency terms. These frequency terms are used in order to describe our frequency-based descriptors as follows:

$$F = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_5 \end{bmatrix} \tag{10}$$

where the vector $\mathbf{f}_i = [d_1, d_2, \dots, d_n]^T$ is an n -dimensional vector, which describes the frequency coefficients output from FFT for each position trajectory of the end effectors and n denotes the number of frames.

3.4. Motion Classification

To learn the affective state features, we use the SVM learning algorithm, which is a state-of-the-art classifier technique for multi-class classification. There are several SVM multi-class classification strategies: one-against-one, one-against-rest, and the directed acyclic graph SVM (DAG-SVM) [19]. Among these alternatives, the one-against-one method trains the classifier by choosing every different two classes such that there will be $n * (n - 1) / 2$ possible binary subclassifiers for a problem of n classes. For prediction of a new motion sequence, every subclassifier judges its category, and the category voted for highest score is selected as the class of the test data. On the other hand, the one-against-rest method uses a two-class SVM and compares every given class with all the others put together such that n subclassifiers are constructed. DAG-SVM method separates train data into two category and goes on separation in each category until only two classes are left. Because we use four affective state classes for classification in this work, we use the one-against-one method; however, for

Table 1. Confusion matrices for classification with different feature vectors.

	Sad	Happy	Angry	Relaxed	Sad	Happy	Angry	Relaxed	Sad	Happy	Angry	Relaxed	Sad	Happy	Angry	Relaxed
Sad	100.0	0.0	0.0	0.0	85.71	7.14	0.0	7.14	78.57	7.14	0.0	14.28	85.71	7.14	0.0	7.14
Happy	14.28	71.42	0.0	14.28	14.28	71.42	7.14	7.14	7.14	78.57	0.0	14.28	7.14	78.57	7.14	7.14
Angry	0.0	14.28	85.71	0.0	0.0	7.14	92.85	0.0	0.0	7.14	0.0	7.14	0.0	7.14	92.85	0.0
Relaxed	7.14	14.28	0.0	78.57	7.14	14.28	0.0	78.57	7.14	0.0	7.14	85.71	7.14	14.28	0.0	78.57
		Frequency				Posture				Dynamic				Frequency+Posture		
Sad	85.71	7.14	0.0	7.14	92.85	7.14	0.0	0.0	92.85	7.14	0.0	0.0	0.0	0.0	0.0	0.0
Happy	7.14	78.57	0.0	14.28	7.14	85.71	0.0	7.14	7.14	85.71	0.0	7.14	7.14	85.71	0.0	7.14
Angry	0.0	7.14	85.71	7.14	0.0	7.14	92.85	0.0	0.0	7.14	0.0	7.14	0.0	7.14	92.85	0.0
Relaxed	7.14	7.14	0.0	85.71	0.0	7.14	7.14	85.71	7.14	0.0	0.0	92.85	7.14	0.0	0.0	92.85
		Frequency+Dynamic				Posture+Dynamic				Frequency+Dynamic				Frequency+Dynamic		

larger set of classes, the one-against-rest and DAG-SVM can also be used.

In linear SVM, given a training data set D with m input vectors $\mathbf{x}_1, \dots, \mathbf{x}_m$ and their corresponding labels y_1, \dots, y_m , where $y_i \in -1, +1$ and x_i is a p -dimensional real vector, the class of a new input data is determined by evaluating the sign of $y(x)$:

$$y(x) = w^T \phi(\mathbf{x}) + b \tag{11}$$

It also should be noted that, although in linear SVM, the decision boundary is $w^T x + b$, in nonlinear SVM, the radial basis function (RBF) is

$$y(x) = w^T e^{-\gamma \|\mathbf{x}\|^2} + b \tag{12}$$

where $\phi(x) = e^{-\gamma \|\mathbf{x}\|^2}$ defines the Gaussian RBF.

Basically, three main issues need to be considered during support vector classification: feature selection, kernel model selection, and the selection of the parameters of the kernel function. Because in this work we have more than two classes (sad, happy, angry, and relaxed), multi-class

SVM is used with the corresponding class labels: 0 (sad), 1 (happy), 2 (angry), 3 (relaxed). In our case, the input feature vectors are constructed by combining feature sets from different feature categories. There are three feature sets:

$$\mathcal{P} = \begin{bmatrix} E \\ T \\ B \\ C \end{bmatrix}, \mathcal{D} = \begin{bmatrix} V \\ A \\ J \end{bmatrix}, \mathcal{F} = [F]. \tag{13}$$

for posture feature set, dynamic feature set, and frequency-based feature set, respectively.

3.4.1. Model Selection.

As discussed earlier, SVM presents several different functions in order to map data to higher dimensional spaces. These kernel function are linear, polynomial, RBF and sigmoid. Whereas there are two parameters for an RBF kernel, C and γ , linear kernel has only a penalty parameter. In most cases, the RBF kernel is reasonable first choice. Best combination of C and γ can be determined by a grid search with exponentially growing sequences of C and γ , for example, $C \in \{2^{-5}, 2^{-3}, \dots, 2^{13}, 2^{15}\}$; $\gamma \in$

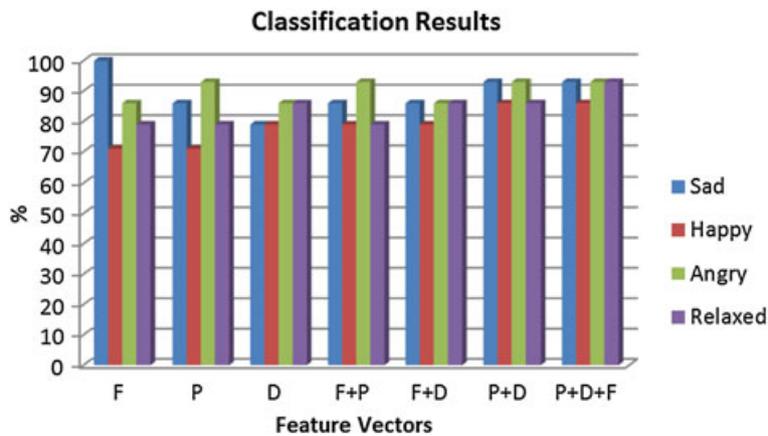


Figure 4. Classification results.

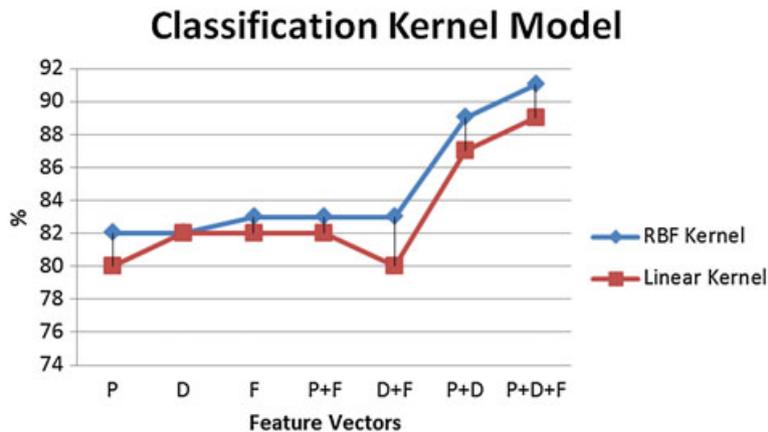


Figure 5. Classification results.

$\{2^{-15}, 2^{-13}, \dots, 2^1, 2^3\}$. In our study, we conducted pilot tests where we compare different values, and the results showed that the RBF kernel yields better performance and accuracy than the other kernels for this application.

3.4.2. Model Validation.

Our classifier was tested with leave-one-out cross validation to evaluate the accuracy of our model. During validation, we selected a single example motion sequence from the database as the validation sample, and the remaining clips were as the training data. We repeated this procedure until each example motion in the database has been used once as the validation sample. The prediction of the classifier is considered to be true if the predicted class label by classifier matches the class label of the test motion data.

4. RESULTS

In this section, the experimental results for several combinations of features that are explained in Section 3 are analyzed in terms of their classification accuracy. We designed seven experiments to evaluate the accuracy of features. In the first three experiments, the frequency-based feature set \mathcal{F} , posture feature set \mathcal{P} and dynamic feature set \mathcal{D} were tested separately. For the following three experiments, these feature sets were paired together in a vector form of $[\mathcal{F}, \mathcal{P}]$, $[\mathcal{F}, \mathcal{D}]$ and $[\mathcal{P}, \mathcal{D}]$. In the seventh and last experiment, all features were combined together resulting in a vector $[\mathcal{F}, \mathcal{P}, \mathcal{D}]$.

For the evaluation of the experiments, confusion matrices [20] are presented that contain information about actual and predicted classification carried out in the classification process. They are also used for evaluating the performance of the classification. Table 1 reports the confusion matrix for each experimental condition. In each section of the table, the values on the diagonal represent the correct recognition rates, whereas the remaining entries gives the rates of incorrect hits (classifications).

Figure 4 shows the results of the style-based classification of human motion graphically. The overall success of the system is 91% where the feature vector is the combination of three feature sets $[\mathcal{F}, \mathcal{P}, \mathcal{D}]$ used in experiment 7. The experiment results show that the choice of features has a significant impact on the accuracy of prediction.

Because determining the best kernel to use in classification is important, we also compared the classification results of the linear kernel to the RBF kernel. Figure 5 shows that the RBF kernel yields better prediction results than the linear kernel for classification. However, a careful parameter search of the kernel parameters was necessary. Especially, for the RBF kernel parameters, we used $C = 1000$ and $\gamma = 0.0001$.

5. CONCLUSION

On the basis of the motivation that the style of an actor is conveyed through emotion, the study presented here is an

attempt to address the question whether there are emotion-specific indicators in human motion and whether they can be extracted and quantified into some measures in terms of affective state descriptors. We conducted a comprehensive analysis of these descriptors (features) that fall into three different categories—posture descriptors, dynamic descriptors, and frequency-based descriptors, and compared their prediction accuracy. According to the experimental results, the most accurate results are achieved when all the descriptors are combined. With a limited set of emotional states (sad, happy, angry, and relaxed) that represent each quadrant of Russell's circumplex model, the prediction accuracy was as high as 91%.

There are several limitations of the system, which we plan to address in the future work. First, we selected only four representative emotional states from the more comprehensive set of emotions, and the method would be beneficial with more states. Secondly, although our classifier was accurate, it would be necessary to compare its performance with the classification answers by real users. Third, we have investigated the walking motion, for a more accurate unification of the training set; however, more general types of motions need to be investigated. Furthermore, it would be beneficial to investigate other types of affective descriptors, particularly semantic ones. Finally, future work needs to address motion synthesis with the use of the learned affective state descriptors.

ACKNOWLEDGEMENTS

This work is supported by the Scientific and Technical Research Council of Turkey (TUBITAK, project number 112E105). We would like to thank all those who participated in the experiments for this study.

REFERENCES

1. Russell JA. A circumplex model of affect. *Journal of Personality and Social Psychology* 1980; **39**: 1161–1178.
2. Etemad SA, Arya A. 3d human action recognition and style transformation using resilient backpropagation neural networks, In *IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009. ICIS 2009*, Vol. 4, Ottawa, ON, Canada, 2009; 296–301.
3. Ali S, Basharat A, Shah M. Chaotic invariants for human action recognition, In *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007*, Rio de Janeiro, Brazil, 2007; 1–8.
4. Huimin Q, Yaobin M, Wenbo X, Zhiqian W. Recognition of human activities using SVM multi-class classifier. *Pattern Recognition Letters* 2010; **31**: 100–111.

5. Mori T, Nejigane Y, Shimosaka M, Segawa Y, Harada T, Sato T. Online recognition and segmentation for time-series motion with hmm and conceptual relation of actions, 2005.
6. Parameswaran V, Chellappa R. View invariants for human action recognition, In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003*, Vol. 2, College Park, MD, USA, 2006; II-613-19.
7. Troje NF. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision* 2002; **2**(5): 371-387.
8. Coulson M. Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior* 2004; **28**(2): 117-139.
9. Wallbott HG. Bodily expression of emotion. *European Journal of Social Psychology* 1998; **28**: 879-896.
10. Atkinson AP, Dittrich WH, Gemmell AJ, Young AW. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 2004; **33**: 717-746.
11. Kapur A, Kapur A, Virji-Babul N, Tzanetakis G, Driessen PF. Gesture-based affective computing on motion capture data. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*. Springer Verlag, Berlin, Heidelberg, 2005; 1-7.
12. Rose C, Cohen MF, Bodenheimer B. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications* 1998; **18**(5): 32-40.
13. Grochow K, Martin SL, Hertzmann A, Popović Z. Style-based inverse kinematics. *ACM Transactions on Graphics* 2004; **23**: 522-531.
14. Ma W, Xia S, Hodgins JK, Yang X, Li C, Wang Z. Modeling style and variation in human motion, In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Beijing, China, 2010; 21-30.
15. Senin P. Dynamic Time Warping Algorithm Review. 2, Department of Information and Computer Sciences, University of Hawaii, Honolulu, Hawaii 96822, Dec. 2008.
16. Heloir A, Courty N, Gibet S, Multon F. Temporal alignment of communicative gesture sequences: Research articles. *Computer Animation and Virtual Worlds* 2006; **17**(3-4): 347-357.
17. Krüger B, Tautges J, Weber A, Zinke A. Fast local and global similarity searches in large motion capture databases. In *2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, Aire-la-Ville, Switzerland, 2010; 1-10.
18. De Silva PR, Bianchi-Berthouze N. Modeling human affective postures: an information theoretic characterization of posture features: Research articles. *Computer Animation and Virtual Worlds* 2004; **15**: 269-276.
19. Hsu CF, Lin CJ. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks* 2002; **13**: 415-425.
20. Stehman SV. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 1997; **62**(1): 77-89.

AUTHORS' BIOGRAPHIES



graphics, and motion analysis and synthesis.

Gokcen Cimen received her BS in Computer Science from Izmir Institute of Technology (Turkey) in 2010, and she is currently a final year MS at Bilkent University (Turkey) for Computer Graphics and Animation. Her current research interests include data-driven character animation, computer



Hacer Ilhan completed her MS degree at the Mathematics Program at Hacettepe University (Turkey) in 2011. She is currently pursuing her PhD degree at Hacettepe University. Her research interests include applied mathematics, computer graphics, and data driven character animation.



His research interests include networked virtual environments, mobile graphics, computer animation, and human-computer interaction.

Tolga Capin is an assistant professor at the Department of Computer Engineering at Bilkent University. He has received his PhD at the Ecole Polytechnique Federale de Lausanne, Switzerland, in 1998. He has more than 30 journal papers and book chapters, 50 conference papers, and a book.



interests include computer animation, visualization, and computational geometry.

Hasmet Gurcay is currently the director of Computer Graphics Department at Hacettepe University, TR. He received his PhD degree in Mathematics in 1991 from Hacettepe University. Several times, he was a postdoctoral fellow at the University Tubingen in the WSI-GRIS group. His research